Back-end of big data analysis

Academic Year: (2019 / 2020) **Review date: 04-10-2018**

Department assigned to the subject: Computer Science and Engineering Department

Coordinating teacher: CALLE GOMEZ, FRANCISCO JAVIER

Type: Compulsory ECTS Credits : 3.0

Year : 1 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

- Programming skills
- Experience in Windows environment

Yet not required, it is also desirable to meet:
- basics of structured databases (at least, Relational Model)
- notions of relational algebra (or at least, set theory)
- basics of data languages (specially SQL)
- basics of Javascript

All of these will be introduced in class, and further materials provided so the student can catch up (or at least, get introduced) at home.

OBJECTIVES

This course aims are
- To distinguish different approaches to secondary storage, depending upon the needs, but focusing on storage with analytic purposes.
- To learn the wide diversity of solutions, and specifically to get introduced to some of the most widespread tools supporting Big Data implantations.

Agenda will cover from the information acquisition and preparation to the manipulation on some DBMS
Therefore, it is largely characterized by a practical focus

To achieve these goals, the student must acquire a set of generic capabilities, knowledge, skills and attitudes.

Cross/Generic Capabilities
- o Analysis and synthesis abilities
- o Organize and plan abilities
- o Troubleshooting
- o Ability to apply knowledge in practice

Specific Capabilities
Cognitive (Knowledge): Storage paradigms, Information Lifecycle, back-end solutions for Big Data
Procedural/Instrumental (Know how)
- Data manipulation
- Information acquisition and preparation
- Use of a document oriented DBMS
- Use of a column oriented DBMS

Attitudinal (To be):
ability to design queries (creativity), concerns about the effectiveness and the efficiency, and ability to discuss and clarify the diverse solutions to each specific problem

DESCRIPTION OF CONTENTS: PROGRAMME

ITEM 1. Storage Paradigms
- Introduction to Storage: archives and files
- Databases and DBMS
- Evolution of Storage and DBMS: OLTP vs. OLAP
- Massive storage: ROLAP vs. RTOLAP

ITEM 2. Structured Storages
- Relational databases
- Analytical Processing on Relational DB

ITEM 3. Information Acquisition and preparation
- Acquisition and extraction
- Transformation, Cleaning & Integration

ITEM 4. Supporting Big Data: document-oriented approach and MongoDB

ITEM 5. Supporting Big Data: key-value / column approach and Cassandra

LEARNING ACTIVITIES AND METHODOLOGY

Lectures:
Highly practical lectures, most will take place at a computer room.
Most tools involved in the course are free, so the student is encouraged to practice with them (after classes).
It is an ambitious (and large) syllabus: some further details will be left for the student to review at home (complementary lectures).

Assignments:
Except from first item (introductory and of theoretical nature) every block will end up with an assignment, for the student to solve at home.
Practical exercises (similar to those in the assignment) will be proposed in the classroom.

Tutoring sessions:
Individual tutoring sessions will be provided for solving eventual doubts, so the student can catch up with the group in case some parts are not found to be completely understood.

ASSESSMENT SYSTEM

Four main blocks (corresponding to items 2, 3, 4 and 5) will end up with an assignment to be solved by students individually at home. First one is optional (the student can score 10 over 10 without submitting it, but can find some useful extra score by solving it).

Final test (8th week) to assess global learning.

Global mark will be calculated as the average of four compulsory scores
(first assignment is taken into account for increasing final test grade, up to 40% of its score).

% end-of-term-examination:	25
% of continuous assessment (assigments, laboratory, practicals…):	75

BASIC BIBLIOGRAPHY

- J. Calle Course Teaching Materials (provided via Aula Global webpage) Each item will have specific references (mostly, links to webpages where documentation on tools usage, syntax, etc. can be freely accessed), Aula Global, 2018

ADDITIONAL BIBLIOGRAPHY

- Elmasri, R. y Navathe, SB Fundamentals of Database Systems. , Pearson .
- Hurwitz, J, Nugent, A, Halper, F, Kaufman, M. Big data for dummies , Wiley, 2013
- Ramakrishnan, R. y Gehrke, J. Database management systems. , McGraw Hill. .
- Warden P. Big Data glossary. A guide to the New Generation of data tools. , O¿Reilly , 2011
- null Express Learning: Database Management Systems. , ITL Education Solutions Lt. Pearson India Pubs, 2012
- Rijmenam, M.V. Think Bigger (ISBN-13: 978-0-8144-3415-4), Amacom, 2014